

LINKAGE DISEQUILIBRIUM WITH THE ISLAND MODEL¹

TOMOKO OHTA

National Institute of Genetics, Mishima, 411, Japan

Manuscript received October 19, 1981

Revised copy accepted February 5, 1982

Linkage disequilibrium between two linked loci was studied for a finite population with a subdivided population structure. Wright's island model was used; extinction and replacement of colonies were also incorporated. Two alleles (A_1 and A_2 at the first locus, and B_1 and B_2 at the second locus) with symmetric mutation rates were assumed, and equilibrium properties of linkage disequilibrium coefficients were analyzed. In terms of analogy with the subdivision of inbreeding coefficient, the variance of linkage disequilibrium is divided into several components: D_{IS}^2 (variance of within-colony disequilibrium), D_{ST}^2 (variance of correlation of A_1 and B_1 of different gametes of one colony relative to that of the total population), and D_{IT}^2 (total variance of disequilibrium). Other subdivisions are D'_{IS} (variance of correlation of A_1 and B_1 of one gamete of a colony relative to that of the average gamete of the population) and D'_{ST} (variance of the ordinary disequilibrium of the whole population). When migration is limited, the variance becomes large if the correlation of A_1 and B_1 of one colony is taken relative to that of the whole population (D_{ST}^2 and D'_{IS}). Also, when the rate of extinction-replacement of colonies is high, the whole-population disequilibrium coefficient (D'_{ST}) can become fairly large. Observed linkage disequilibria, such as those among markers in the major histocompatibility complex of man and mouse, may well be explained by limited migration, without assuming epistatic natural selection.

LINKAGE disequilibrium, or nonrandom association of alleles at two loci, has been studied extensively both theoretically and experimentally (see HEDRICK, JAIN and HOLDEN 1978 for review). Two causes are considered to be mainly responsible for the nonrandom association of alleles: epistatic natural selection and random genetic drift. Perhaps the most impressive example of observed linkage disequilibrium is that among loci of the major histocompatibility complex (MHC) of man and mouse (BODMER 1979, KLEIN and FIGUEROA 1981). In interpreting such an observed fact, these authors propose that functionally inter-related genes are clustered to form a "supergene"; that is, they consider the observed linkage disequilibrium to be caused by epistatic fitness interactions, *i.e.*, natural selection.

One flaw in the theoretical studies is that they are primarily concerned with random mating populations (see KIMURA and OHTA 1971); under such an assumption, linkage disequilibrium due to random genetic drift is generally thought

Contribution No. 1414 from the National Institute of Genetics, Mishima, Shizuoka-ken, 411, Japan

insignificant. This is probably the reason that the observed linkage disequilibrium, such as in MHC, is usually attributed to natural selection. However, the principal species studied, mouse and man, have (or have had) structured populations; and migration between the subpopulations (colonies) may be quite limited. Then, recombination between the loci in question becomes less effective (because of fewer heterozygotes) and linkage disequilibrium may be created among the local colonies. The purpose of this report is to clarify the magnitude of linkage disequilibrium due to random drift when the population is subdivided. We shall show that strong linkage disequilibrium may occur between the colonies when migration is limited.

BASIC THEORY

Because of the great complexity of the problem, we will present a very simplified situation, the Wright's island model (WRIGHT 1940) is used. Let us consider a population (species) consisting of n colonies (subpopulations or demes), where each colony is assumed to consist of N breeding individuals (effective size). Let m be the migration rate among the colonies per generation. Note that, in the island model, every colony receives in one generation a fraction m of its individuals from the entire population. In addition, I assume, as did MARUYAMA and KIMURA (1980), that each colony is subject to extinction at rate λ per generation and that, whenever a colony is extinct, it is immediately replaced by a line derived from individuals from a single colony in the population.

Again, for simplicity, a two-allele model with symmetric mutation is assumed:

$$A_1 \xrightleftharpoons[v]{v} A_2 \quad \text{and} \quad B_1 \xrightleftharpoons[v]{v} B_2,$$

and c is the recombination rate between the two loci. Table 1 summarizes the evolutionary processes and the parameters.

Under this model, how much nonrandom association is expected between the A and B loci? Clearly, the expected value (mean) of the linkage disequilibrium coefficient is zero at equilibrium. Thus, in the following, the variance of the linkage disequilibrium coefficient is investigated. The variance is created by

TABLE 1

Processes and their parameters for the two-locus model

Recombination :	Rate = $c/\text{generation}$
Random genetic drift within a colony :	effective population size = N
Migration (island model)	Rate = $m/\text{generation}$
Number of colonies	n
Extinction-replacement of colonies	Rate = $\lambda/\text{generation}$
Mutation (symmetric, two-allele)	Rate = $v/\text{generation}$
	$\alpha = 2(m + \lambda)/n$
Variables of the i -th colony	
g_{1i} = frequency (A_1B_1), g_{2i} = frequency (A_1B_2), g_{3i} = frequency (A_2B_1) and g_{4i} = frequency (A_2B_2)	
$x_i = g_{1i} + g_{2i}$ and $y_i = g_{1i} + g_{3i}$	
Variables of the total population	
$G_j = \sum_{i=1}^n g_{ji}/n$ for $j = 1 \sim 4$, $X = \sum_{i=1}^n x_i/n$ and $Y = \sum_{i=1}^n y_i/n$	

random genetic drift, and in our model it is pronounced in each colony. Particularly when migration is limited, random drift in each colony would prevail and different types of chromosomes would spread in different colonies, increasing the variance of linkage disequilibrium.

Let us consider the i -th colony, and let g_{1i} , g_{2i} , g_{3i} and g_{4i} be the frequencies of A_1B_1 , A_1B_2 , A_2B_1 and A_2B_2 chromosomes in the i -th colony. Let $x_i = g_{1i} + g_{2i}$ and $y_i = g_{3i} + g_{4i}$ be the frequencies of A_1 and B_1 in the i -th colony. Further, let us denote by large letters the population average of these variables: $G_j = \sum_{i=1}^n g_{ji}/n$ for $j = 1 \sim 4$, $X = \sum_{i=1}^n x_i/n$ and $Y = \sum_{i=1}^n y_i/n$. These notations of variables are also summarized in Table 1. The linkage disequilibrium coefficient may be defined at various levels; the within-colony disequilibrium coefficient is, for the i -th colony,

$$\begin{aligned}\delta_i &= g_{1i}g_{4i} - g_{2i}g_{3i} \\ &= g_{1i} - x_i y_i .\end{aligned}\quad (1)$$

Disequilibrium coefficients between the colonies, or the correlation of A_1 and B_1 within a colony relative to that of the total population are defined in the following ways:

$$d_i = g_{1i} - XY \quad (2)$$

and

$$d_i' = x_i y_i - XY \quad (2')$$

$$d_i'' = g_{1i} - G_1 . \quad (2'')$$

Finally, the disequilibrium coefficient of the total population may be expressed,

$$D = G_1 - XY . \quad (3)$$

The above subdivision of linkage disequilibrium coefficient is analogous to that of inbreeding coefficients of WRIGHT (1940) in structured populations (see also NEI 1975, HARTL 1981). A still closer analogy may be found in two-locus properties of various mating schemes (COCKERHAM and WEIR 1977) and of transmission genetics of mitochondria (OHATA 1981).

The expected values of the above disequilibrium coefficients become zero at equilibrium

$$E\{\hat{\delta}_i\} = E\{\hat{d}_i\} = E\{\hat{d}_i'\} = E\{\hat{d}_i''\} = E\{\hat{D}\} = 0,$$

where the caret (\wedge) means equilibrium and E denotes the expectation. Thus, in the following I shall formulate the change of the moments necessary for the variance of disequilibrium coefficients; and hence I shall evaluate the magnitude of the variances at equilibrium.

To calculate variances, one needs moments of g_{1i} , x_i , y_i , G_1 , X and Y up to the fourth. Let us start from the second moments. Let $s_1 = E\{x_i^2\}$ and $s_2 = E\{X^2\}$. Then, as in MARUYAMA and KIMURA (1980), it is possible to get transition formulae of s_1 and s_2 from one generation to the next. The basic theory is quite similar to the calculation of identity coefficients of gene members of multigene families (see OHATA 1980a for review). The identity coefficient within one chro-

mosome corresponds to s_1 , and that between different chromosomes to s_2 . The analogy also extends to the model of extranuclear genes (TAKAHATA and MARUYAMA 1981), and TAKAHATA (1982) has started to formulate linkage disequilibrium for such genes. Now $s_1 (= E\{x_i^2\})$ changes, through random genetic drift within a colony, by the amount $E\{x_i(1-x_i)\}/2N$, in one generation. Through migration, it changes by the amount, $2m\{E(X^2) - E(x_i^2)\}$, and, by mutation, $2v\{E(x_i) - 2E(x_i^2)\}$, per generation. Therefore, we have for the change of s_1 in one generation, by assuming the parameters, $1/(2N)$, m , $v \ll 1$,

$$\Delta s_1 = \frac{1}{2N} \{E(x_i) - E(x_i^2)\} + 2m\{E(X^2) - E(x_i^2)\} + 2v\{E(x_i) - 2E(x_i^2)\} . \quad (4)$$

As for $s_2 = E\{X^2\}$, it does not change by random drift within a colony, but is influenced by the extinction-replacement process of colonies. In one generation, a colony has the probability λ of going extinct and being replaced, and s_2 changes by the amount $\frac{2\lambda}{n}\{E(x_i^2) - E(X^2)\}$ when n is sufficiently large (see MARUYAMA and KIMURA 1980). Similarly, it changes by migration, with the amount $\frac{2m}{n}\{E(x_i^2) - E(X^2)\}$. Its change by mutation is similar to that of s_1 , $2v\{E(X) - 2E(X^2)\}$. Therefore we have

$$\Delta s_2 = \frac{2(\lambda + m)}{n} \{E(x_i^2) - E(X^2)\} + 2v\{E(X) - 2E(X^2)\} . \quad (5)$$

At equilibrium, $\Delta s_1 = \Delta s_2 = 0$, and $E(\hat{x}_i) = E(\hat{X}) = 1/2$. By replacing these values in the equations (4) and (5), one gets the equilibrium second moments.

$$\hat{s}_1 = E\{\hat{x}_i^2\} = \frac{1 + 4N\left\{v + \frac{nmv}{\lambda + m + 2nv}\right\}}{2 + 16N\left\{v + \frac{nmv}{\lambda + m + 2nv}\right\}} \quad (6)$$

and

$$\hat{s}_2 = E\{\hat{X}^2\} = \frac{2(\lambda + m)\hat{s}_1 + nv}{2(\lambda + m) + 4nv} \quad (7)$$

where a caret (\wedge) means an equilibrium value. Note that, because of the symmetry of the model, we have

$$E\{\hat{y}_i^2\} = \hat{s}_1 \text{ and } E\{\hat{Y}^2\} = \hat{s}_2 . \quad (8)$$

Let us proceed to the third moments. Let \mathbf{t} be the vector of moments, $\mathbf{t} = (t_1, t_2, t_3, \dots, t_6)$ with

$$\begin{aligned} t_1 &= E\{g_{1i}x_i\} \\ t_2 &= E\{x_i^2y_i\} \\ t_3 &= E\{G_1X\} \\ t_4 &= E\{x_iy_iX\} \\ t_5 &= E\{X^2Y\} \\ t_6 &= E\{x_i^2Y\} . \end{aligned} \quad (9)$$

and

$$\Delta \mathbf{t} = \begin{bmatrix} -(6\nu + \frac{1}{2N} + c + 2m) & c & 2m & 0 & 0 & 0 \\ \frac{1}{N} & -(6\nu + \frac{3}{2N} + 3m) & 0 & 2m & 0 & m \\ \alpha & 0 & -(6\nu + c + \alpha) & c & 0 & 0 \\ 0 & \frac{1}{2N} & \frac{1}{2N} & -(6\nu + \frac{1}{2N} + \alpha + 2m) & 2m & 0 \\ 0 & \alpha & 0 & 2\alpha & -(6\nu + 3\alpha) & \alpha \\ 0 & 0 & 0 & 0 & 2m & -(6\nu + \frac{1}{2N} + 2m + \alpha) \end{bmatrix} + \begin{bmatrix} vE(g_{ii} + x_i y_i + x_i^2) + \frac{1}{2N} E(g_{ii}) \\ vE(2x_i y_i + x_i^2) + \frac{1}{2N} E(x_i y_i) \\ vE(G_i + XY + X^2) \\ vE(x_i y_i + XY + X^2) \\ vE(2XY + X^2) \\ vE(2XY + x_i^2) + \frac{1}{2N} E(XY) \end{bmatrix} \quad (10)$$

Again the changes of \mathbf{t} by various forces in one generation are obtained. Some of the derivations will be given in the APPENDIX. It becomes, by assuming the parameters, $1/(2N)$, c , m , λ and $v \ll 1$.

where $\alpha = 2(\lambda + m)/n$. At equilibrium, it can be shown that

$$\hat{t}_1 = \hat{t}_2 = \hat{t}_6 = \hat{s}_1/2 \text{ and } \hat{t}_3 = \hat{t}_4 = \hat{t}_5 = \hat{s}_2/2. \quad (11)$$

Also, from the symmetry of the model, we have $E\{x_i \gamma_i^2\} = \hat{t}_2$, $E\{G_1 Y\} = \hat{t}_3$ and so on at equilibrium.

The fourth moments are even more complicated. Let \mathbf{f} be the vector of moments, $\mathbf{f} = (f_1, f_2, \dots, f_{13})$ such that,

$$\begin{aligned} f_1 &= E\{g_{1i}^2\} & f_8 &= E\{(x_i^2)(\gamma_i^2)\} \\ f_2 &= E\{G_1^2\} & f_9 &= E\{(x_i \gamma_i)(x_j \gamma_j)\} \\ f_3 &= E\{g_{1i} x_i \gamma_i\} & f_{10} &= E\{(x_i^2 \gamma_i) Y\} \\ f_4 &= E\{G_1 (x_i \gamma_i)\} & f_{11} &= E\{(x_i^2) Y^2\} \\ f_5 &= E\{(g_{1i} x_i) Y\} & f_{12} &= E\{(x_i \gamma_i) X Y\} \\ f_6 &= E\{G_1 X Y\} & f_{13} &= E\{X^2 Y^2\} \\ f_7 &= E\{x_i^2 \gamma_i^2\} \end{aligned} \quad \text{and} \quad (12)$$

The changes of the above moments in one generation are expressed by the following equation. Some examples of their derivations will be given in the APPENDIX.

As before, the equilibrium values are obtained by putting $\Delta \mathbf{f} = 0$.

In terms of the above moments, the variances of linkage disequilibrium coefficients take the following form when random drift, migration, mutation, recombination and extinction-replacement of colonies balance each other. By means of the analogy with subdivision of inbreeding coefficients of WRIGHT (1940) (see also NEI 1975 and HARTL 1981), the subscripts *IS*, *ST* and *IT*, are used. Note that the process of the frequency change is stochastic because of random genetic drift within a colony and of extinction-replacement of colonies. In the following formulation, expectation is taken for distribution and the symbol E denotes expectation with respect to such distributions of x_i , γ_i and g_{1i} . The variance component of disequilibrium within a colony is denoted as D_{IS}^2 and it becomes

$$D_{IS}^2 = E\{\delta_i^2\} = E\{g_{1i}^2 - 2g_{1i}x_i\gamma_i + x_i^2\gamma_i^2\} = \hat{f}_1 + \hat{f}_7 - 2\hat{f}_3. \quad (14)$$

The variance of correlation of A_1 and B_1 of different gametes of one colony relative to that of the total population is

$$D_{ST}^2 = E\{d_i^2\} = E\{x_i^2\gamma_i^2 - 2x_i\gamma_iXY + X^2Y^2\} = \hat{f}_7 + \hat{f}_{13} - 2\hat{f}_{12}. \quad (15)$$

The variance of correlation of A_1 and B_1 of one gamete in a colony relative to that of the total population is

$$D'_{IS}{}^2 = E\{d_i''^2\} = E\{g_{1i}^2 - G_1^2\} = \hat{f}_1 - \hat{f}_2. \quad (16)$$

The next component is the variance of disequilibrium of the total population:

$$D_{ST}^2 = E\{D^2\} = E\{G_1^2 - 2G_1XY + X^2Y^2\} = \hat{f}_2 + \hat{f}_{13} - 2\hat{f}_6. \quad (17)$$

Finally, the total variance of linkage disequilibrium becomes

$$D_{IT}^2 = E\{d_i^2\} = E\{g_{1i}^2 - 2g_{1i}XY + X^2Y^2\} = \hat{f}_1 + \hat{f}_{13} - 2\hat{f}_6. \quad (18)$$

Note that the additivity principle as in inbreeding coefficients does not hold for the above components of linkage disequilibrium, and we have

$$D_{IT}^2 \neq D_{IS}^2 + D_{ST}^2. \quad (19)$$

However, it does hold for the other subdivision:

$$D_{IT}^2 = D_{IS}^2 + D_{ST}^2. \quad (20)$$

A useful measure of linkage disequilibrium may be the standardized value, standardized linkage disequilibrium, sometimes called squared standard linkage deviation (HILL and ROBERTSON 1968, OHTA and KIMURA 1969, SVED and FELDMAN 1973). Standardization is appropriate for D_{IS}^2 and D_{ST}^2 .

$$\sigma_{IS}^2 = \frac{E\{\delta_i^2\}}{E\{x_i y_i (1-x_i)(1-y_i)\}} = \frac{\hat{f}_1 + \hat{f}_7 - 2\hat{f}_3}{0.25 - \hat{s}_1 + \hat{f}_7} \quad (21)$$

and

$$\sigma_{ST}^2 = \frac{E\{D^2\}}{E\{XY(1-X)(1-Y)\}} = \frac{\hat{f}_2 + \hat{f}_{13} - 2\hat{f}_6}{0.25 - \hat{s}_2 + \hat{f}_{13}}.$$

Standardized linkage disequilibrium is only slightly influenced by mutation rate, and it takes a similar value even in the infinite allele model of mutation, as my further analysis shows, so long as other parameters remain the same (OHTA 1982).

DISCUSSION AND EXAMPLES

This study is the first theoretical evaluation of the magnitude of linkage disequilibrium for a structured population at equilibrium. In particular, the theory for subdividing the variance of disequilibrium into components has been established. Many attempts have been made to investigate the effect of limited migration on linkage disequilibrium (MAYNARD SMITH 1968, CAVALLI-SFORZA and BODMER 1971, SINNOCK and SING 1972, NEI and LI 1973, PROUT 1973, KARLIN 1973, LI and NEI 1974, FELDMAN and CHRISTIANSEN 1975, SLATKIN 1975). However, all these studies treat special transient cases, not a general situation as given here. Also, the effect of nonrandom mating such as selfing has been shown to reduce effective recombination and therefore to increase disequilibrium (KARLIN 1969; WEIR and COCKERHAM 1969, 1973; CLEGG, ALLARD and KAHLER 1972; WEIR and HILL 1980). In these studies, however, no population subdivision was taken into account.

BROWN and FELDMAN (1981) have developed a series of indices on the variance of the number of heterozygous loci for measuring multilocus associations of structured populations. Their variance components include some with two loci, which may be directly obtained from the moments of equation (12). The present theory provides a theoretical basis for any measure on allelic association of two loci for subdivided populations.

By using the theory developed in the previous section, the variances of linkage disequilibrium coefficients are numerically obtained for some interesting cases. Figure 1 shows three measures of linkage disequilibrium, D_{IS}^2 , D_{IT}^2 and $D_{ST}^{\prime 2}$ as functions of migration rate. In the figure, migration rate, m , changes from zero to 0.015. Other parameters are n (number of colonies) = 200, N (colony size) = 100, v (mutation rate) = 10^{-4} , c (recombination rate) = 0.005 and λ (extinction-replacement rate) = 10^{-2} . Such a population structure is considered to be realistic for mammalian species such as the mouse or ancient human populations. From the figure, it can be seen that, when migration is limited, the total variance of linkage disequilibrium becomes large. For example, when $m = 0.001$ ($Nm = 0.1$), meaning effective migration of one individual in 10 generations, $D_{IT}^2 = 0.106$; when $m = 0.005$ ($Nm = 0.5$), $D_{IT}^2 = 0.052$. In these examples, recombination rate is 0.01, and therefore such a large linkage disequilibrium is expected between loci one centimorgan apart.

Next, let us examine how disequilibrium is influenced when the recombination value changes. Figure 2 illustrates three measures of variance of disequilibrium as functions of recombination rate. Parameters are; $c = 0 \sim 0.01$, $n = 200$, $N = 100$, $v = 10^{-4}$, $\lambda = 10^{-2}$ and $m = 0.01$. Again large variance of linkage disequilibrium appears between the colonies. When c becomes larger, the variance gets smaller; however, the rate of decrease is rather small.

As shown above, the variance of disequilibrium may get fairly large when the population is structured, if disequilibrium within a colony is measured relative to that of the total population (D_{ST}^2 and D_{IS}^2). This is because different types of chromosomes spread in different colonies by random genetic drift, and migration is insufficient to mix chromosome types. It is generally said that, when $4Nm > 1$, genetic differentiation is negligible among the colonies (WRIGHT 1969). However, this statement does not apply for linked loci. Linkage and limited migration apparently enhance the effects of each other.

The variance of total-population disequilibrium ($D_{ST}^{\prime 2}$) is not very large in the above examples; however, when extinction-replacement of colonies is rapid and the number of colonies is small, it may become significant. Particularly, when standardized linkage disequilibrium ($\sigma_{ST}^{\prime 2}$) is measured, this tendency becomes clear. Table 2 shows some examples of when the number of colonies is small or extinction rate is high.

To compare linkage disequilibrium due to random drift with that due to natural selection, I include Figure 3, D^2 expected from epistatic natural selection. KIMURA's (1956a) result is used. In his model, it is assumed that the first locus (A_1, A_2) is maintained polymorphic by overdominance, whereas, for the second locus, the first allele (B_1) is advantageous in combination with A_1 but

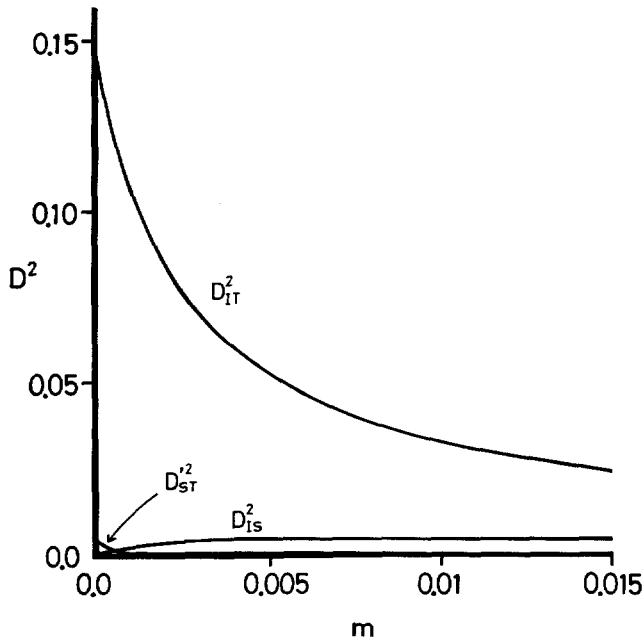


FIGURE 1.—Variances of three measures of linkage disequilibrium as functions of migration rate (m), at equilibrium. Parameters are, $N = 100$, $v = 10^{-4}$, $n = 200$, and $c = 0.005$, $\lambda = 0.01$. They are numerically calculated, by generating the third and fourth moments (equations 10 and 13) from one generation to the next for at least $2/v$ generations.

disadvantageous in combination with B_1 and the second allele (B_2) is in reverse relationship. The selection scheme is shown in Table 3, where t is the coefficient of overdominance at the A locus, and ε is the intensity of epistasis at the B locus. In this model, the frequencies of A_1B_1 and A_2B_2 chromosomes increase and D is expected to be positive. KIMURA (1956a) has shown that, at equilibrium, D becomes

$$\hat{D} = \frac{1}{2} \left(\sqrt{\frac{1}{4} + \beta^2} - \beta \right), \quad (22)$$

where $\beta = (1+t)c/\varepsilon$. When $t \ll 1$, β is roughly the ratio of recombination fraction and epistatic selection intensity. Also, the positive value of D is stable when $c < (t^2 - \varepsilon^2)/\{4t(1+t)\}$ (KIMURA 1956b). For a comparison with disequilibrium due to random drift, the relationship between \hat{D}^2 and β is given in Figure 3. To maintain a large disequilibrium, β has to be much less than unity, as can be seen from the figure; in other words, the recombination value has to be much less than the epistatic selection intensity.

From the above analyses, one predicts that the relationships, $D_{IS}^2 > D_{ST}^2$ and $D_{ST}^2 > D_{IS}^2$, usually hold when migration is limited. On the other hand, if epistatic natural selection is responsible for linkage disequilibrium, as in the model of KIMURA (1956), one predicts $D_{IS}^2 < D_{ST}^2$ and $D_{ST}^2 < D_{IS}^2$. Therefore,

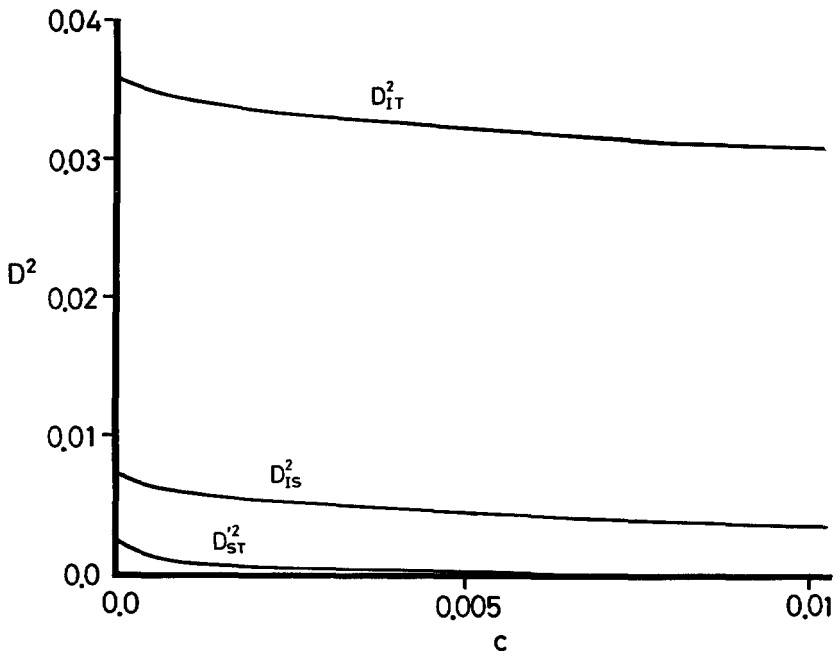


FIGURE 2.—Variances of three measures of linkage disequilibrium as functions of recombination rate (c), when various forces balance each other. Parameters are, $N = 100$, $\nu = 10^{-4}$, $n = 200$ and $m = \lambda = 0.01$.

TABLE 2

*Variances of three measures of linkage disequilibrium when rate of extinction-replacement of colonies (λ) is high and number of colonies (n) is small.**

Parameters $N = 20$, $\nu = 10^{-4}$, $m = 10^{-3}$, $c = 0.005$		Second moments		Linkage disequilibrium				
		s_1	s_2	D_{IS}^2	σ_{IS}^2	D_{IT}^2	D_{ST}^2	σ'_{ST}^2
$n = 100$	$\lambda = 0.0$	0.479	0.261	0.0011	0.731	0.1645	0.0007	0.012
	$\lambda = 0.02$	0.487	0.371	0.0004	0.723	0.0994	0.0026	0.151
	$\lambda = 0.04$	0.490	0.411	0.0002	0.704	0.0706	0.0018	0.212
	$\lambda = 0.06$	0.491	0.432	0.0001	0.682	0.0547	0.0012	0.245
	$\lambda = 0.08$	0.492	0.444	0.0001	0.660	0.0447	0.0009	0.265
	$\lambda = 0.1$	0.493	0.453	0.0001	0.638	0.0377	0.0007	0.276
$\lambda = 0.04$	$n = 40$	0.493	0.453	0.0001	0.637	0.0373	0.0007	0.277
	$n = 60$	0.492	0.437	0.0001	0.674	0.0505	0.0011	0.254
	$n = 80$	0.491	0.423	0.0002	0.693	0.0614	0.0014	0.232
	$n = 120$	0.489	0.401	0.0003	0.711	0.0784	0.0020	0.196

* Note fairly large values of σ'_{ST}^2 (standardized linkage disequilibrium of the whole population).

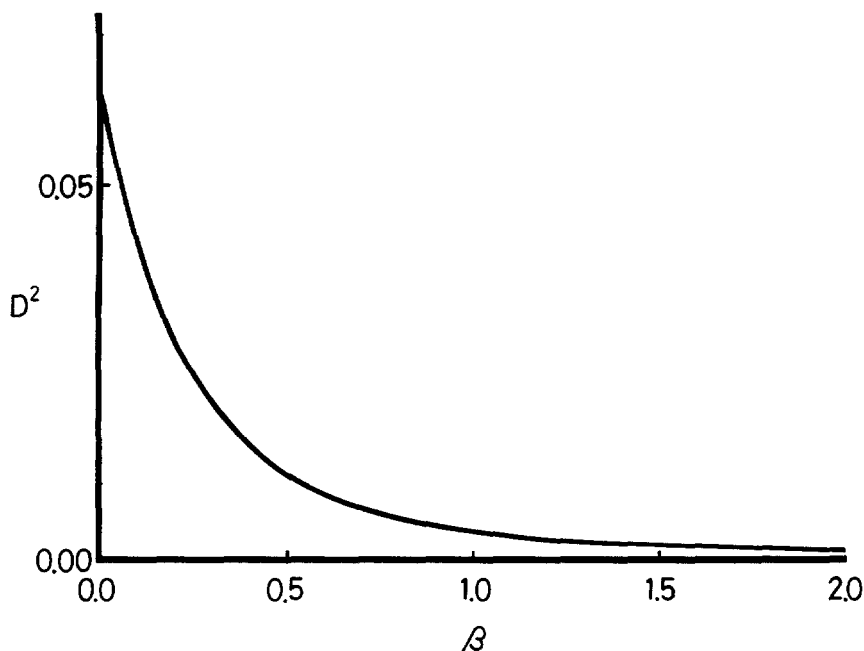


FIGURE 3.—Linkage disequilibrium due to epistatic natural selection. For comparison with that due to random genetic drift, D^2 is shown. Abscissa is $\beta = (1 + t)c/\epsilon$ where t is heterozygote advantage at the first locus, c is recombination rate and ϵ is epistatic selection coefficients.

the ratio, D_{IS}^2/D_{ST}^2 or D_{ST}^2/D_{IS}^2 , may be a measure for testing which of the factors, epistasis or population subdivision, is the main cause of observed linkage disequilibrium. The data as given in KLEIN and FIGUEROA (1981) would be useful for such a test.

Even selectionists now realize that subdivided population structure may be partly responsible for observed linkage disequilibrium (BODMER and BODMER 1978, WILLS 1981). The question is, which is more important, limited migration or epistasis? I suggest that tight linkage and limited migration are the main cause of disequilibrium rather than natural selection, for the following reasons: Although some MHC markers are known to have a strong correlation with certain autoimmune diseases (*e.g.*, SASAZUKI, McDEVITT and GRUMET 1977), such diseases would not have much influence on fitness. In particular, epistatic selection coefficients much larger than the recombination rate would be difficult to

TABLE 3

*Fitnesses of the model of epistatic natural selection**

	A_1A_1	A_1A_2	A_2A_2
B_1B_1	$1 + \epsilon$	$1 + t$	$1 - \epsilon$
B_1B_2	1	$1 + t$	1
B_2B_2	$1 - \epsilon$	$1 + t$	$1 + \epsilon$

* From KIMURA (1956a).

imagine even for tightly linked markers. Also, as pointed out before, mammalian species have structured populations with limited migration among the colonies (KLEIN 1979). In addition, because of the rapid development of modern human society, present races, such as Caucasian or Oriental populations, may be mixtures of previously separated colonies. Then large D_{ST}^2 or D'_{IS}^2 would be reflected by a large disequilibrium within the present populations.

On the other hand, invertebrates such as *Drosophila* may have different population structures from those of mammals. Data on single loci already suggest less differentiation in *Drosophila* species than in mammals (AYALA *et al.* 1974). Studies of linkage disequilibrium suggest that there is a very small amount of disequilibrium with the exception of loci involved in chromosome inversions (MUKAI, METTLER and CHIGUSA 1971; MUKAI, WATANABE and YAMAGUCHI 1974; LANGLEY, TOBARI and KOJIMA 1974; PRAKASH 1974; ZOUROS *et al.* 1974). Thus, it would be reasonable to conclude that migration is sufficient in most *Drosophila* species to mix the various genotypes among the local colonies.

Data on genetic polymorphisms are now available not only for the higher organisms mentioned above, but also for *E. coli*. SELANDER and LEVIN (1980) surveyed electrophoretic variation at 20 enzyme loci in 109 clones of *E. coli* from natural populations. A most notable fact is that linkage disequilibrium is pronounced among the clones. In this case, there is linkage disequilibrium between all pairs of loci in the genome, indicating that the sexual recombination is very rare in this organism. Therefore, the recombination rate $c \approx 0$ in our model, and a large variance of disequilibrium is expected.

As discussed above, the present theory provides a useful tool for understanding observed linkage disequilibrium. However, the model may be too simple; classical two-allele mutation and island model migration are assumed. It would be desirable to extend this work to a more general mutation scheme as has been done for panmictic populations (OHTA and KIMURA 1971, HILL 1975, GRIFFITH 1981, TAKAHATA 1981), and I have recently worked out this problem for the infinite allele model (OHTA 1982). It would also be preferable to study linkage disequilibrium with a more general population structure.

A final remark is concerned with the model where random drift takes place at more than one level. In our case, it occurs within the colony by random genetic drift and in the whole population by extinction-replacement of colonies. In the model of multigene families (OHTA 1980a), random drift occurs on the chromosome by unequal crossing over and in the population through random sampling of gametes. To clarify linkage disequilibrium between amino acid sites within the variable region of immunoglobulins, I performed the following approximate analyses (OHTA 1980b). Because of the much lower rate of drift in the chromosome than in the population, and because of sufficient exchange of genes between the chromosomes, gene diversity measured by amino acid identity is almost the same whether it is measured within the chromosome or between the chromosomes. Likewise, I assumed that the variance of linkage disequilibrium between amino acid sites would be similar whether it is measured within the chromosome or between chromosomes. By means of the present theory, it is now possible to examine this assumption. Table 4 shows some examples comparable to the case

TABLE 4

*Variances of three measures of linkage disequilibrium when random drift within the colony is slow compared to the high rate of extinction-replacement of colonies**

Parameters $n = 20, m = \lambda = 0.1$		Second moments		Linkage disequilibrium					
		s_1	s_2	D_{IS}^2	σ_{IS}^2	D_{IT}^2	D_{ST}^2	σ_{ST}^2	$D_{IS}'^2$
c									
$N = 500$ $v = 10^{-4}$	0.0			0.0037	0.090	0.0043	0.0036	0.086	0.0006
	0.0005	0.2970	0.2961	0.0020	0.048	0.0025	0.0018	0.044	0.0006
	0.001			0.0014	0.034	0.0019	0.0012	0.029	0.0006
	0.0025			0.0008	0.019	0.0013	0.0006	0.015	0.0006
$N = 10^3$ $v = 10^{-5}$	0.0			0.0034	0.241	0.0036	0.0034	0.239	0.0002
	0.0005	0.3831	0.3828	0.0005	0.040	0.0008	0.0005	0.037	0.0002
	0.001			0.0003	0.023	0.0005	0.0003	0.020	0.0002
	0.0025			0.0002	0.011	0.0004	0.0001	0.009	0.0002

* Note that the three disequilibrium coefficients are almost the same.

of this multigene family. The approximation seems to work well. In other words, the variances of three measures of linkage disequilibrium are almost equal ($D_{IS}^2 \approx D_{ST}^2 \approx D_{IT}^2$) and $D_{IS}'^2 = E\{(x_i y_i - XY)^2\} \approx 0$, indicating that there is little differentiation among the colonies with respect to the two loci. In such cases, linkage disequilibrium is expected to be similar whether it is measured in the colony (within a chromosome in case of a multigene family) or between colonies (between the chromosomes in the multigene model).

I thank M. KIMURA for his continuous encouragement throughout the course of this study. I also thank K. MORIWAKI and N. TAKAHATA for their stimulating discussions, and J. F. CROW, W. HILL and B. S. WEIR for their many valuable suggestions and comments on the manuscript, which greatly improved the presentation.

LITERATURE CITED

- AYALA, F. J., M. L. TRACEY, L. G. BARR, J. F. McDONALD and S. PÉREZ-SALAS, 1974 Genetic variation in natural populations of five *Drosophila* species and the hypothesis of the selective neutrality of protein polymorphisms. *Genetics* **77**: 343-384.
- BODMER, W. F., 1979 Gene clusters and the HLA system. pp. 205-229. In: *Human Genetics: Possibilities and Realities*, Ciba Foundation Series **66**. Elsevier, North-Holland.
- BODMER, W. F. and J. G. BODMER, 1978 Evolution and function of the HLA system. *British Medical Bull.* **34**: 309-316.
- BROWN, A. H. D. and M. W. FELDMAN, 1981 Population structure of multilocus associations. *Proc. Nat. Acad. Sci. U.S.A.* **78**: 5913-5916.
- CAVALLI-SPORZA, L. L. and W. F. BODMER, 1971 *The Genetics of Human Populations*. W. H. Freeman, San Francisco.
- CLEGG, M. T., R. W. ALLARD and A. L. KAHLER, 1972 Is the gene the unit of selection? Evidence from two experimental plant populations. *Proc. Nat. Acad. Sci. U.S.A.* **69**: 2474-2478.
- COCKERHAM, C. C. and B. S. WEIR, 1977 Digenic descent measures for finite populations. *Genet. Res.* **30**: 121-147.
- FELDMAN, M. W. and F. B. CHRISTIANSEN, 1975 The effect of population subdivision on two loci without selection. *Genet. Res.* **24**: 151-162.

- GRIFFITH, R. C., 1981 Neutral two-locus multiple allele models with recombination. *Theor. Pop. Biol.* **19**: 169-186.
- HARTL, D. L., 1981 *A Primer of Population Genetics*, Sinauer Associates Books, Addison-Wesley, New York.
- HEDRICK, P., S. JAIN and L. HOLDEN, 1978 *Multilocus Systems in Evolution*, *Evolutionary Biology*, Vol. 11. Edited by M. K. HECHT, W. C. STEERE and B. WALLACE. Plenum, New York.
- HILL, W. G., 1975 Linkage disequilibrium among multiple neutral alleles produced by mutation in finite population. *Theor. Pop. Biol.* **8**: 117-126.
- HILL, W. G. and A. ROBERTSON, 1968 Linkage disequilibrium in finite populations. *Theor. Appl. Genetics* **38**: 226-231.
- KARLIN, S., 1969 *Equilibrium Behavior of Population Genetics Models with Non-random Mating*. Gordon and Breach, London. ———, 1973 Sex and infinity: a mathematical analysis of the advantages and disadvantages of genetic recombinations. pp. 155-194. In: *The Mathematical Theory of the Dynamics of Biological Populations*. Edited by M. S. BARTLETT and R. W. HIORNS. Academic Press, London.
- KIMURA, M., 1956a A model of a genetic system which leads to closer linkage by natural selection. *Evolution* **10**: 278-287. ———, 1956b Rules for testing stability of a selective polymorphism. *Proc. Nat. Acad. Sci. U.S.A.* **42**: 336-340.
- KIMURA, M. and T. OHTA, 1971 *Theoretical Aspects of Population Genetics*. Princeton Univ. Press, Princeton, N. J.
- KLEIN, J., 1979 The major histocompatibility complex of the mouse. *Science* **203**: 516-521.
- KLEIN, J. and F. FIGUEROA, 1981 Polymorphism of the mouse *H-2* loci. *Immunol. Rev.* **60**: 23-57.
- LANGLEY, C. H., Y. N. TOBARI and K. KOJIMA, 1974 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics* **78**: 921-936.
- LI, W. H. and M. NEI, 1974 Stable linkage disequilibrium without epistasis in subdivided populations. *Theor. Pop. Biol.* **6**: 173-183.
- MARUYAMA, T. and K. KIMURA, 1980 Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proc. Nat. Acad. Sci. U.S.A.* **77**: 6710-6714.
- MAYNARD SMITH, J., 1968 Evolution in sexual and asexual populations. *Amer. Nat.* **102**: 469-473.
- MUKAI, T., L. E. METTLER and S. I. CHIGUSA, 1971 Linkage disequilibrium in a local population of *Drosophila melanogaster*. *Proc. Nat. Acad. Sci. U.S.A.* **68**: 1065-1069.
- MUKAI, T., T. K. WATANABE and O. YAMAGUCHI, 1974 The genetic structure of natural populations of *Drosophila melanogaster*. XII. Linkage disequilibrium in a large local population. *Genetics* **77**: 771-793.
- NEI, M., 1975 *Molecular Population Genetics and Evolution*. North-Holland/American Elsevier, Amsterdam, Oxford, New York.
- NEI, M. and W. H. LI, 1973 Linkage disequilibrium in subdivided populations. *Genetics* **75**: 213-219.
- OHTA, T., 1980a *Evolution and Variation of Multigene Families*. Lecture Notes in Biomathematics, Vol. 37. Springer, New York. ———, 1980b Linkage disequilibrium between amino acid sites in immunoglobulin genes and other multigene families. *Genet. Res.* **36**: 181-197. ———, 1981 Two-locus problems in transmission genetics of mitochondria and chloroplasts. *Genetics* **96**: 543-555. ———, 1982 Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proc. Nat. Acad. Sci. U.S.A.* **79**: 1940-1944.

- OHTA, T. and M. KIMURA, 1969 Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics* **63**: 229-238. ———, 1971 Linkage disequilibrium between two segregating nucleotide sites under steady flux of mutations in a finite population. *Genetics* **68**: 571-580.
- PRAKASH, S., 1974 Gene differences between the sex ratio and standard gene arrangements of the X chromosome and linkage disequilibrium between loci in the standard gene arrangement of *Drosophila pseudoobscura*. *Genetics* **77**: 795-804.
- PROUT, T., 1973 Appendix to MITTON, J. B. and R. C. KOEHN, Population genetics of marine pelecypods, III. Epistasis between functionally related isoenzymes in *Mytilus edulus*. *Genetics* **73**: 487-496.
- SASAZUKI, T., H. O. McDEVITT and F. C. GRUMET, 1977 The association between genes in the major histocompatibility complex and disease susceptibility. *Ann. Rev. Med.* **28**: 425-452.
- SELANDER, R. K. and B. R. LEVIN, 1980 Genetic diversity and structure in *Escherichia coli* populations. *Science* **210**: 545-547.
- SINNOCK, P. and C. F. SING, 1972 Analysis of multilocus genetic systems in Tecumseh, Michigan, II. Consideration of correlation between non-alleles in gametes. *Am. J. Hum. Genet.* **24**: 393-415.
- SLATKIN, M., 1975 Gene flow and selection in a two locus system. *Genetics* **81**: 787-802.
- SVED, J. A. and M. W. FELDMAN, 1973 Correlation and methods for one and two loci. *Theor. Pop. Biol.* **4**: 129-132.
- TAKAHATA, N., 1981 Linkage disequilibrium, genetic distance and evolutionary distance under a general model for genes or a part of genome. *Genet. Res.* **39**: 63-67. ———, Linkage disequilibrium of extranuclear genes under neutral mutations and random genetic drift. Submitted to *Theor. Pop. Biol.*
- TAKAHATA, N. and T. MARUYAMA, 1981 A mathematical model of extranuclear genes and the genetic variability maintained in a finite population. *Genet. Res.* **37**: 291-302.
- WEIR, B. S. and C. C. COCKERHAM, 1969 Group inbreeding with two linked loci. *Genetics* **63**: 711-742. ———, 1973 Mixed self and random mating at two loci. *Genet. Res.* **21**: 247-252.
- WEIR, B. S. and W. G. HILL, 1980 Effect of mating structure on variation in linkage disequilibrium. *Genetics* **95**: 477-488.
- WILLS, C., 1981 *Genetic Variability*. Oxford University Press, New York.
- WRIGHT, S., 1940 Breeding structure of populations in relation to speciation. *Amer. Nat.* **74**: 232-248. ———, 1969 *Evolution and the Genetics of Populations*. Vol. 2. *The Theory of Gene Frequencies*. The Univ. of Chicago Press, Chicago and London.
- ZOUROS, E., C. B. KRIMBAS, S. TSAKAS and M. LOUKAS, 1974 Genic versus chromosomal variation in natural populations of *Drosophila subobscura*. *Genetics* **78**: 1223-1244.

Corresponding editor: B. WEIR

APPENDIX

Derivation of coefficients for the change of the third and fourth moments:

There is no space to present derivations of all coefficients of the equations (10) and (13), therefore only a couple of equations are derived below. All calculations in the following are the one-generation changes of moments. Let us start with $t_1 = E\{g_{1i}x_i\}$. Through mutations, t_1 changes by the amount,

$$\begin{aligned}\Delta_{\text{mut}}\{t_1\} &= E\{v g_{1i}(1 - 2x_i) + v x_i(x_i + \gamma_i - 4g_{1i})\} \\ &= vE\{g_{1i} + x_i^2 + x_i\gamma_i - 6g_{1i}x_i\},\end{aligned}\tag{A1}$$

where $\Delta_{\text{mut}}\{A\}$ denotes expected change of A by mutation. The change of moments by random

genetic drift within a colony may be obtained by using the diffusion equation method of OHTA and KIMURA (1969). The amount of change of t_1 becomes, denoting the expected change by random drift as Δ_{drift} ,

$$\Delta_{\text{drift}}\{t_1\} = E\left\{\frac{1}{2N} g_{1i}(1 - x_i)\right\} . \quad (\text{A2})$$

Next, by recombination with rate c , t_1 changes by the amount

$$\Delta_{\text{rec}}\{t_1\} = E\{c(x_i^2 y_i - g_{1i} x_i)\} \quad (\text{A3})$$

where $\Delta_{\text{rec}}\{\cdot\}$ denotes the expected change by recombination. Finally through migration with rate m ,

$$\begin{aligned} \Delta_{\text{mig}}\{t_1\} &= E\{m x_i(G_1 - g_{1i}) + m g_{1i}(X - x_i)\} \\ &= 2mE\{G_1 X - g_{1i} x_i\} , \end{aligned} \quad (\text{A4})$$

where $\Delta_{\text{mig}}\{\cdot\}$ represents expected change by migration. On the average, t_1 is not influenced by extinction-replacement of colonies. Thus, sum of the above equations (A1)–(A4) gives the coefficients of the first line of equation (10).

Next, two moments, f_1 and f_{13} , are chosen for illustration. It is easy to obtain the changes in f_1 by mutation and recombination.

$$\Delta_{\text{mut}}E\{f_1\} = 2vE\{g_{1i}(x_i + y_i - 4g_{1i})\} \quad (\text{A5})$$

and

$$\Delta_{\text{rec}}E\{f_1\} = cE\{g_{1i}(x_i y_i - g_{1i})\} . \quad (\text{A6})$$

Through migration,

$$\Delta_{\text{mig}}E\{f_1\} = 2mE\{g_{1i}(G_1 - g_{1i})\} = 2mE\{G_1^2 - g_{1i}^2\} . \quad (\text{A7})$$

Transformation by random genetic drift within the colony is again calculated by the method of OHTA and KIMURA (1969). It becomes

$$\Delta_{\text{drift}}E\{f_1\} = \frac{1}{2N} E\{g_{1i} - g_{1i}^2\} . \quad (\text{A8})$$

The total amount of Δf_1 may be obtained by summing (A5)–(A8), since f_1 is not influenced by extinction-replacement of colonies. On the contrary, $f_{13} = E\{X^2 Y^2\}$ changes through extinction-replacement process, but neither by random genetic drift within the colony, nor by recombination. In calculating, consider the following relationship:

$$\Delta\{E(X^2 Y^2)\} = E\{2X^2 Y(\Delta Y) + 2XY^2(\Delta X) + X^2(\Delta Y^2) + Y^2(\Delta X^2) + 4XY(\Delta XY)\} . \quad (\text{A9})$$

Now, ΔX and ΔY through migration and extinction-replacement process are zero on the average, and ΔX^2 , ΔY^2 and ΔXY have the coefficient, $\alpha = 2(m + \lambda)/n$, such that this fraction of the second moment with respect to the total population are replaced by those moments within the colony (MARUYAMA and KIMURA 1980). Therefore, we have,

$$\Delta_{\text{mig-ext}}\{E(X^2)\} = \alpha E\{x_i^2 - X^2\} , \quad (\text{A10})$$

where $\Delta_{\text{mig-ext}}\{\cdot\}$ denotes expected change by migration and extinction-replacement process, and ΔY^2 and ΔXY have similar expressions. By mutation, f_{13} changes by the following amount:

$$\Delta_{\text{mut}}\{f_{13}\} = 2vE\{X^2 Y(1 - 2Y) + XY^2(1 - 2X)\} . \quad (\text{A11})$$

Therefore, we have, for the total change of f_{13} in one generation,

$$\Delta\{f_{13}\} = 2vE\{X^2 Y + XY^2 - 4X^2 Y^2\} + \alpha E\{x_i^2 Y^2 + X^2 y_i^2 + 4x_i y_i XY - 6X^2 Y^2\} . \quad (\text{A12})$$

Note that, because of symmetry of the model, and since we are interested in the magnitude of linkage disequilibrium when various forces balance each other, symmetric relationships, such as $E(X^2 Y) = E(Y^2 X)$ or $E(x_i^2 Y^2) = E(X^2 y_i^2)$, are assumed in formulating equation (13). Thus, the equation is inappropriate for generating moments in certain transient situations.